DOCUMENT RESUME

ED 262 068 TM 850 544

AUTHOR Howes, Anthony C.
TITLE Evaluating the Validity of Chapter 1 Data: Taking a Closer Look.
PUB DATE Apr 85
NOTE 29p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Check Lists; Data Collection; Elementary Education; *Evaluation Methods; *Program Evaluation; *Quality Control; *Standardized Tests; *State Departments of Education; State Programs; Technical Assistance; Testing Problems; Test Norms
IDENTIFIERS *Education Consolidation Improvement Act Chapter 1; Oklahoma State Department of Education; Technical Assistance Centers

ABSTRACT

At the request of the Oklahoma Department of Education, the Technical Assistance Center 3 West conducted a statewide review of local Chapter 1 evaluation data to provide possible explanations for the unusual standardized test score gains or losses reported by a number of local school districts. During the two months of quality control visits, 35 districts were interviewed. A checklist, which is appended, was designed to record the different types of quality control errors that were found in each district. The six most frequently occurring errors are described: (1) use of outdated tests; (2) students tested outside norm dates; (3) students tested at non-equivalent periods of time; (4) incorrect norm tables used; (5) greater than 25% rate of chance-level scores; and (6) use of improper out-of-level testing procedures. The implications of and suggestions for overcoming the detected errors are presented. As a result of this study, policy changes were implemented at the State Education Agency level which altered testing practices, data collection, and aggregation procedures throughout the state. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED262068

Evaluating the Validity of Chapter 1 Data: Taking a Closer Look

Written By

Anthony C. Howes

TAC Program Director

Powell Associates, Inc.

Presented By

JoAnn Canales

Evaluation Specialist

Powell Associates, Inc.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Howes, Anthony C.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☑ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Paper presented at the annual meeting of the American Educational Research Association, Chicago, Illinois, March 31-April 4, 1985.

TM 850 544

Evaluating the Validity of
Chapter 1 Data: A Model for Taking
a Closer Look

ANTHONY C. HOWES, Powell Associates, Inc./
Technical Assistance Center--Region 3-West
JOANN CANALES, Powell Associates, Inc./
Technical Assistance Center--Region 3-West

Anthony C. Howes
Powell Associates, Inc.
3355 Bee Cave Road, Suite 305
Austin, Texas 78746

The impact of the Technical Assistance Center services in Oklahoma over the past funding period has focused on "Quality Control" issues. The TAC reviewed the annual evaluation results of 55 school districts in Oklahoma which were selected to participate as a result of reported excessive NCE gains or losses. After reviewing the data from individual sites, TAC staff prepared a report for the SEA summarizing the types of errors found and their consequences. As a result of this study, policy changes were implemented at the SEA level which altered testing practices and data collection and aggregation procedures throughout the state.

EVALUATING THE VALIDITY OF CHAPTER 1 DATA: TAKING A CLOSER LOOK

INTRODUCTION

Historically school districts have been required to evaluate Chapter 1 programs in terms of a pre-post test analysis of standardized test scores. The success or failure of programs is often judged by the magnitude of NCE gains or losses. For NCE scores to be used in evaluating the success of programs, it is imperative that State Education Agencies (SEA) and Local Education Agencies (LEA) become very knowledgeable of testing practices, data collection procedures, and aggregation procedures to help insure the validity of evaluation results. Any judgement regarding the success of programs that is based on NCE scores must be predicated on valid test data.

At the request of the Oklahoma Department of Education, the staff of TAC 3 West became involved in a statewide review of local Chapter 1 evaluation data to help provide possible explanations for the unusual gains or losses reported by a number of districts. Based on the 1981-1982 State Chapter 1 evaluation

report, districts were identified for participation in this study.

Although LEAs were encouraged to participate, involvement was strictly voluntary. The districts participating in the study were selected on the basis of the following criteria:

1. NCE losses at three or more grade levels in reading, math, or a combination of both with losses equaling or exceeding six NCEs at each of the three grade levels; or

2. NCE gains at three or more grade levels in reading, math, or a combination of both with gains equaling or exceeding twenty NCEs at each of the three grade levels.

3. Norm Referenced tests were used to evaluate the Chapter 1 program

During a series of meetings scheduled in February, March, and April 1984 throughout Oklahoma, SEA and TAC staff met with representatives from each participating LEA. At that time, the district's evaluation process and data were reviewed systematically to determine testing and computation procedures.

A number of conditions were discovered as a result of these consultations; however, one general statement could summarize the findings of the study. The rules established by the test publishers must be followed when using a standardized achievement test for program evaluation. Deviating from the publisher's

rules may significantly alter the outcome of the evaluation results. Test publishers establish test administrative and scoring procedures to help districts obtain meaningful test results. Ignoring the rules can seriously affect the ability to make accurate interpretations of the test data.

RESULTS

Thirty districts were interviewed over the two month period of time that the quality control visits were conducted. Although visits were scheduled with an additional 15 districts, (45 total) several came to the central meeting location without the proper data and approximately 10 cancelled just prior to the meetings. The meetings typically lasted from 1.5 to 2 hours per individual school district. Of the 30 districts participating in the study, 22 were implementing a fall-spring testing interval to evaluate their Chapter 1 programs. The remaining 8 were on a spring to spring testing cycle.

A checklist was designed to record the different types of quality control errors that were found to be occuring in each district. A total of 15 possible errors were included on the checklist. Of the 15 errors 13 were found to be occuring in the districts with varying degrees of frequency. In addition, several other types of errors not included on the checklist were found to occur, but

were very isolated in their frequency of occurence. Appendix A reflects the particular checklist used in this study.

Table 1 reveals the six most frequently occuring quality control errors found during the investigation. The table also indicates the frequency of errors occuring in each of fall-spring and spring-spring testing intervals. In addition the data is examined for all districts participating in the study regardless of the testing interval.

Each of the 6 errors were found to occur in 13 percent or more of all districts participating in the study. Only one of the six errors was found to occur less than 13 percent of the time for either fall-spring or spring-spring testing intervals. Of the districts participating, 22 were on a fall-spring testing interval and 8 were implementing a spring-spring interval.

A brief description of the 6 most frequently occuring quality control errors is presented to provide a common frame of reference for the participants.

1. Use of Outdated Tests. A norm-referenced test is usually regarded as outdated when the norms have not been revised for a period of time that exceeds 6 years. It is important to make a distinction between developing new items for a test and developing new norms. Test publishers frequently re-norm tests without creating new items. This is a widely accepted practice among psychometricans. However, tests

should be re-normed every 6 years to establish new patterns of student performance.

2. Tested Outside Norm Dates. Test publishers usually test students over a period of one month when developing national norms for a norm-referenced test. The midpoint of the testing period is commonly referred to as the empirical norm date. Typically empirical norm dates are available for fall, spring, and in some instances mid-year. It is generally recommended that districts test within two weeks before or after the empirical norm date.

3. Tested at Non-equivalent Periods of Time. Districts may choose to test several weeks earlier or later than the empirical norm date. When testing occurs earlier or later at pre test, the post test should be adjusted accordingly to keep the period of time between the pre and post test the same as the test publisher.

4. Incorrect Norm Tables Used. Test publishers norms tables differ based on the level of the test, the form of the test, the particular subtest administered, the date of testing, and the particular score conversion tables needed. It is very important that the proper norms tables be used in all data analysis and score conversions.

5. Greater Than 25% Rate of Chance-Level Scores. A certain number of items are generally answered correctly on

multiple choice tests regardless of whether the student knows the correct response. The probability of a student making the correct response when randomly guessing at test items is directly related to the number of choices the student has per item. On the average the mathematical probability of guessing the correct answer on a true-false question is 50 percent; while the probability drops to 25 percent for four choice items. For example, on the average a student would be expected to guess the correct response on 10 out of 20 true-false items (50%) and only 5 out of 20 items (25%) on a four choice per item test. A chance-level-score on an achievement test indicates the student taking the test could have received the same score by randomly guessing at the test items. When more than 25 percent of the students in a specific population score at chance the validity of the test results is questioned.

6. Improper Out-of-Level Testing Procedures Used. Districts choosing to test students with a test level other than the test publishers recommended level must adhere to the proper procedures for converting those scores back in level. For example; if a district chooses to test third grade students with a second grade test, the students test scores must be converted back in-level or the test results will be spuriously elevated. Strict adherence to proper conversion procedures is essential for the test results to be valid.

Table 1 indicates that the most frequently occuring quality control error was testing outside of the empirical norm dates. This error was found to occur in 100 percent of the districts on a fall-spring testing schedules, in 38 percent of the districts on spring-spring cycles and 83 percent of all districts participating in the study. It is interesting to note that in Oklahoma this type of error occurs much more frequently with districts on fall-spring (F-S) testing cycles as compared to spring-spring (S-S) cycles. This occured primarily because districts were ignoring the fall empirical norm dates and in many instances were testing the students during the first week of school after the summer vacation.

The second most frequent error found to occur was the use of incorrect norms tables. Fifty-five percent of the districts on S-S testing schedules made these errors, while 63 percent of the districts on S-S schedules were found to make similar errors. A total of 57 percent of all participating districts used incorrect norms tables. In many instances incorrect norms tables were used when districts hand scored their results. In addition, a number of instances occurred where districts used improper norms tables as a result of poor communication by the test publishers. Either empirical norms did not exist or spring norms were developed after the fall norms had been on the market for several years. In the latter situation the district often did not know that spring norms even existed.

The next prevalent error was testing at non-equivalent periods of time. This error occured consistently for both F-S (36%) and S-S (38%) testing cycles for a total of 37 percent of all participating districts. Districts that tested students early in the fall did not adjust the testing schedule in the spring to maintain the same period of time between the pre and post test as did the test publisher. In this instance, adjusting the spring testing dates probably would not have changed the quality of the data because the majority of students were not properly motivated to take the test in the fall after just returning from summer vacation. Similar occurances happened to districts on Sp-Sp schedules; however, adjusting the testing date the following spring would probably be sufficient to counterbalance the effects of early or late testing in the previous year.

Both improper out-of-level testing procedures and a greater than 25 percent rate of chance-level scores occured in 20 percent of all participating districts. Improper out-of-level testing procedures occured consistently between districts on F-S and S-S testing cycles. However, a greater than 25% rate of chance-level scores occured only in districts on F-S intervals. Chance-level scores occur more often in the fall than spring when the same test level is administered at both points. This happens because students normally get more items correct in the spring after an additional 6 to 7 months of instruction has occured between the pre and post test.

The use of outdated tests were found to occur in 13 percent of all participating districts. Outdated tests were used more frequently in districts on S-S testing cycles (38%) than F-S (5%). Each district on a S-S testing schedule used the same outdated test.

CONCLUSIONS

Topic 1--Implications of Using Outdated Tests

When districts use a test with outdated norms they are actually comparing the performance of their students to the performance of students who took the test six or more years ago. More current norms allow for the comparison of students present performance. Over a period of six years student populations change, achievement levels change and test items become obsolete; all of which limits the ability to adequately interpret test results. In addition, test publishers are becoming increasing more sophisticated in the development of tests. Districts who utilize more recent editions will benefit from the better quality tests.

Suggestions for Using Outdated Tests

There are no acceptable procedures for using outdated tests. When tests become outdated a new test should be selected for use

in the district. Periodically districts should go through a test selection process to match a current test with the needs of the school district. No test will meet all the needs of a district, but one can be choosen to meet the majority of those needs.

Topic 2--Implications of Testing Outside Norm Dates

Test publishers establish empirical norm dates to assist school districts in comparing the performance of their students with other students throughout the nation. The basic assumption is that school districts must assess their students during the same period of time the test publisher assessed the norm group.

Generally it is recommended that districts test within two weeks of the emperical norm date. A substantial deviation from the norm date can significantly reduce the validity of the test scores. This is true for both fall and spring testing although deviating from the fall norm date has a greater potential for adversely affecting the quality of the evaluation data.

Districts deviating significantly from the test publisher's fall norm dates have an increased risk of invalidating their results when they test early. In addition to deviating from the norm date they risk testing students who are inadequately prepared to take the test and/or have not developed the proper "mind set" for taking the test so soon after returning from the summer break. Fall test results are usually depressed when compared to the previous spring scores. This phenomenon is frequently referred

to as summer regression or summer drop-off. Testing students early in the fall only adds to the pre existing problem of spuriously low pre test scores which result in inflated gain scores.

Suggestions for Complying with Norm Dates

1. The ideal strategy for complying with norm dates is to test students within two weeks of the publishers empirical norm date.

2. Use a norm-referenced test that has empirical norm dates which coincide with the testing interval used for Chapter 1 evaluation. Older editions of tests may have empirical fall norms but interpolated norms for the spring or fall interpolated norms and spring empirical norms. A district should select a test that has empirical norm dates for the period of time the district anticipates testing its students. This is usually not a problem for more recent editions of norm-referenced tests.

3. Some test publishers have weekly interpolated norms. These norms have been developed to assist districts who can not test within two weeks of the publishers norm date because of scheduling problems. Weekly interpolated norms are only estimates of how students would be expected to perform when tested several weeks prior to or after the actual norm date. Test results are considered most valid when the

districts testing coincides with the publisher's empirical norm dates. However, it is generally accepted that districts may test six weeks prior to or after the publishers norm date if interpolated weekly norms are available.

4. If scoring is done by the publisher, advise them of the week of administration to help insure the appropriate interpolated norms are used to score the test. If scoring is done within the district request interpolated norms from the test publisher and use these when scoring the tests.

Topic 3--Implications of Testing at Non-equivalent Periods of Time

The period of time between the pre and post test for a district's Chapter 1 evaluation should closely approximate the test publishers norm dates. If six months occur between the publisher's fall and spring norms the same period of time should exist between the districts pre and post test. Similarly spring-spring testing should occur at the same time from one year to the next. Shortening or lengthening the period of time between the pre and post test when compared to the publisher's norm dates increases the potential for invalid scores. If the period of time between the districts pre and post test is longer than the test publisher's and interpolated norms are not used the district's gain scores will tend to be artificially inflated.

Conversely, if the period of time between the pre and post test is less than the publisher's the gain scores will tend to be artificially lowered.

Suggestions for Testing at Non-equivalent Periods of Time

If the district chooses to test several weeks early or late at pre test, adjust the post test accordingly to keep the period of time between the pre and post test the same as the test publisher. For example, if the emperical norm dates are October 15 and April 15 and district testing is scheduled for October 1, then spring testing should be moved two weeks forward to April 1. This procedure is recommended if the district is testing early or late at either pre or post test time.

Topic 4--Implications of Using Incorrect Norms Tables

Improper use of the test publisher's norm tables will completely invalidate the district's test results. These errors occur more frequently when districts hand score their test results, although test publisher's on occasion use the incorrect norms tables to score tests.

Districts that inadvertently uses the wrong set of norm tables will spuriously elevate or depress the gain scores depending on whether the incorrect tables were used at pre or post test time. For example, a district on a Sp-Sp testing schedule that uses the previous years spring norms tables instead of the current years

tables to derive percentile scores will falsely elevate the post test scores for the current year and the pre test score for the following year. The end result is that gain scores are overestimated for the current year and underestimated for the following year.

In a similar manner, use of an improper test level or subtest to score a districts test results will result in a misrepresentation of the evaluation data. Using norms from the incorrect form of the test will often result in slight errors in scoring.

SUGGESTIONS FOR USING THE PROPER NORMS TABLES

1. Districts that hand score their test results have a much greater likelihood of using the incorrect norms tables. If your district hand scores be particularly careful when deriving your scores.

2. If your district hand scores their results have a specific set of procedures to be used at each grade level. Make sure these procedures are distributed to each person responsible for hand scoring and deriving the scores.

3. Have another individual double check the work of each individual. Teachers may be responsible for hand scoring

and using look-up tables to convert raw scores into derived scores. They may be given the additional responsibility for checking the work of one of their colleague's who had to go through the same procedure.

4. District's that have their test results machined scored by the district or test publisher should double check their results. Hand scoring a small random sample of the answer sheets and comparing your results with the test publisher's is one simple method for guarding against possible scoring errors.

Topic 5--Implications of Greater Than 25% Chance-Level-Scores

Chance scores often indicate the test is to difficult for the students and/or they are not motivated to take the test. Chance scores may also indicate the student has not been adequately prepared to take the test.

When a large proportion of students score at chance the gain scores are inflated or depressed depending on the cause of the chance-level-scores. For example, if the pre-test is too difficult students tend to randomly guess and will answer more questions correctly than they actually know the answers. Students who randomly guess will score higher when the test level is too difficult. A student may answer 18 of 60 four choice items correctly, but only know the answers for 3 of the questions. By randomly guessing at all but 3 items the student

has answered correctly an additional 15 items. Because the student randomly guessed at the items his or her percentile score will be artificially inflated. When these students are administered an easier level of the test and their scores converted back in-level their derived percentile scores are actually lower when compared to the scores they received on the more difficult test level.

For some students the test level may not be too difficult. Students may score at chance-level because they are not motivated to take the test. Very early testing in the fall is often accompanied by a large percentage of students who score at chance. This often occurs because students have not been prepared for taking the test and do not have the proper mind set to take the test so soon after returning from the summer break. In this situation the chance-level-scores may actually be an underestimate of the students true performance. Testing the student out-of-level will not alter the test results. The students must be sufficiently motivated to take the test to insure valid test results. If a group of students, with average potential, score at chance-level the derived percentile scores will underestimate rather than overestimate their potential.

Gain scores can be artificially inflated or depressed depending on when they occur, what causes them and whether the district is on a F-S or Sp-Sp testing schedule. For those students who score at chance and the test level is too difficult at pre test, the

scores will be overestimated and the gain scores will be underestimated. However if the post test scores are at chance, both these scores and the gain scores will be overestimated. In addition, if the test level is too difficult for the students and the district is on a Sp-Sp testing interval the results will be overestimated for the current year and underestimated for the following year. Similarily if the test level is not too difficult for the students who score at chance, the gain scores can be underestimated or overestimated depending on the testing interval, the year testing took place, and whether the chance-scores occured at pre or post test.

Ideally, good measurement is said to occur when two-thirds of the students answer between one-third and two-thirds of the items correctly. When a fairly large proportion of students are found to be answering correctly the same number of items as would be expected by chance, the validity of the test scores is always questioned. Typically chance level scores are more prevalent in the fall, but regardless of when they occur the ability to interpret student, class, grade, or district data is seriously impaired.

Strategies for Use With Chance Scores

1. The first step in dealing with Chance Scores is knowing how to identify them. If chance scores are known to exist it is then possible to reduce the number of students scoring at chance to reasonable limits. It is impossible to completely eliminate chance scores; however strategies should be undertaken to reduce chance scores.

2. The raw score is usually used in detecting chance scores. After the chance level raw score is figured it is merely a matter of identifying all students who score equal to or lower than the chance score. For example, the chance level raw score for Total Reading in the third grade on Level C of the California Achievement Test is 23. Any student who scores equal to or lower than a raw score of 23 is said to be scoring at chance. The chance-level raw score is different for each test level and subtest score. Tables have been developed to assist in identifying chance scores. If raw scores are not available percentile and NCE scores can usually be used to identify chance scores.

3. Districts having more than 25 percent of its students within a grade level scoring at chance should consider alternatives to reduce the rate of chance scores. In some

instances more student preparation in test-taking skills and motivation may result in a reduction of chance scores. More thorough teacher preparation in giving tests may also improve student performance. Administering an easier in-level test is another procedure frequently used to reduce the rate of chance scores. Finally, out-of-level testing is sometimes conducted to reduce chance scores. Out-of-level testing should never be performed without a complete understanding of the concept and the logistical problems associated with such testing.

Topic 6--Implications Of Using Improper Out-Of-Level Testing Procedures

Many districts knowingly test out-of-level but use improper procedures for converting the scores back in-level. Other districts may not realize they have tested out-of-level and never attempt to convert the scores back in-level. In either case the test results are invalidated. The overall affect is similar to using improper norms tables to convert the raw score into a derived score. Depending on whether out-of-level testing took place one level down or one level up, the scores will be artificially inflated or depressed.

Depending on the particular test used the procedures for converting out-of-level scores back in-level are somewhat different. The basic assumption is to take the out-of-level scale score and convert them into an in-level scale score. The in-level scale score is then converted into a percentile score. Because different test publisher's develop different types of conversion tables it is necessary to examine the specific norm-referenced test to determine the exact procedures.

It is generally recommended that out-of-level testing should not occur more than one level out. That is, third grade students should not be administered a test designed for first grade students. However, third grade students generally could be administered a test designed for second grade students (one level out). Test publisher's generally include some second grade students when norming a test for third graders. This allows the publisher to link different test levels together and to determine how a second grade student will perform on a third grade test. Similarly test publisher's include some fourth grade students when developing norms for the third grade for the same reason.

Strategies To Use When Testing Out-Of-Level

1. Develop a specific set of procedures for out-of-level

testing at each grade level. Identify the proper norms tables to be used depending on whether testing occured in the fall or spring and list their respective page numbers. Identify the proper conversion tables to use when converting the out-of-level scale score into the in-level scale score and identify the proper pages. Identify the proper tables to convert the scale score into the percentile score and cite the page where the tables may be found.

2. Have all conversions double checked by someone who is familiar with the process at that particular grade.

3. Generally do not test out-of-level more than one level. If out-of-level testing occurs at more than one level do not include those scores in your Chapter 1 evaluation.

4. Consult with someone who is very familiar with out-of-level testing if this is an initial attempt.

5. If out-of-level test results are scored by the publisher, double check a random sample of the test protocols. This can be done by scoring the out-of-level tests within the district and converting the scores back in-level. On occasion, test publishers may fail to covert out-of-level scores back in level.

SUMMARY

As a direct outgrowth of the review process in Oklahoma several steps toward more rigorous quality control measures have been taken. First, during regional meetings held in April of 1984 LEAs were encouraged to use a Spring-Spring testing cycle to help avoid the problem of inconvenient fall norms. Districts choosing to test in the fall were required to test within two weeks of the empirical norm date.

In addition, the Oklahoma SEA requested that a summary document be developed by the TAC which outlines the major types of evaluation errors identified in the study. The intent is to distribute the document to all LEAs in the state. The document entitled Oklahoma Quality Control Cookbook will include an explanation of the effects that errors have on the quality of evaluation results. The document will also offer instructions for avoiding these problems in the future.

At the state level the Department of Education will use the information contained in the "Cookbook" to improve state evaluation procedures and to help LEAs improve their planning processes for future Chapter 1 projects.

At the local level, it is hoped that the "Cookbook" will be a valuable tool for school district personnel who have recently assumed responsibility for evaluating Chapter 1 programs.

TABLE 1

FREQUENCY OF QUALITY CONTROL ERRORS

Type of Error	Testing Interval					
	F-S		S-S		Total	
	%	N	%	N	%	N
Use of Outdated Test	5	1	38	3	13	4
Tested Outside Norm Dates	100	22	38	3	83	25
Tested At Non-equivalent Periods of Time	36	8	38	3	37	11
Incorrect Norms Tables Used	55	12	63	5	57	17
Improper Out-of-Level Testing Procedures Used	18	4	25	2	20	6
Greater Than 25% Rate of Chance-Level-Scores	27	6	0	0	20	6
TOTAL N	22		8		30	

APPENDIX

QUALITY CONTROL CHECKLIST

District _____ Contact Person _____

Date _____ Telephone _____

Program _____ Grades _____

Address _____

Yes	No		
___	___	a]	Outdated test used (Norms 10 or more years old)
___	___	b]	Changed test or test edition from pre to post
___	___	c]	Appropriate equating tables used when test changed from pre to post
___	___	d]	Non-matched scores used
___	___	e]	Pretest mean above the 50th NCE
___	___	f]	Tested outside of empirical norm date
___	___	g]	Tested equivalent periods of time around norm dates at both pre- and post-test times
___	___	h]	Incorrect norms tables used
___	___	i]	Averaged inappropriate scores
___	___	j]	Score conversion errors
___	___	k]	Computation errors
___	___	l]	Interpolation errors
___	___	m]	Improper out-of-level testing procedures used
___	___	n]	Tested out-of-level more than one level
___	___	o]	25 percent or greater rate of chance-level scoring

Bibliography

Evaluator's References Title I Evaluation and Reporting System Vol. II, Prepared for U.S. Department of Education Office of Program Evaluation by RMC Research Corporation, Mountain View, California, 1981.

Tallmadge, G. Kasten, Wood, Christine T., and Gamel, Nona N. User's Guide Title I Evaluation and Reporting System, Vol. I, 1981.